

# PENGELOMPOKAN GAMBAR BERDASARKAN WARNA DAN BENTUK MENGUNAKAN FGKA (FAST GENETIC KMEANS ALGORITHM) UNTUK PENCOCOKAN GAMBAR

Farah Zakiah Rahmanti<sup>1</sup>, Entin Martiana K.<sup>2</sup>, S.Kom, M.Kom, Nana Ramadijanti<sup>2</sup>, S.Kom, M.Kom

<sup>1</sup> Mahasiswa, <sup>2</sup> Dosen Pembimbing

Politeknik Elektronika Negeri Surabaya

Institut Teknologi Sepuluh Nopember Kampus ITS Keputih Sukolilo Surabaya 60111, Indonesia

Telp:+62-31-5947280 Fax:+62-31-5946114

Email:far.mizuki@gmail.com

## Abstrak

Kumpulan gambar-gambar digital di berbagai aspek memiliki jumlah yang semakin banyak. Kumpulan gambar tersebut merupakan hasil digitalisasi foto-foto analog, diagram-diagram, lukisan-lukisan, gambar-gambar, dan buku-buku. Cara yang biasa dipakai untuk mencari kumpulan tersebut adalah menggunakan pendekatan pengindeksan dan informasi citra berbasis teks (seperti caption atau keywords). Teknik pencarian gambar seperti ini dinilai tidak praktis karena dua alasan, yakni ukuran basis data yang besar dan subyektif dalam mengartikan gambar.

Berangkat dari hal diatas itulah, dewasa ini telah dikembangkan beragam cara untuk melakukan pencarian gambar yang menggunakan image content suatu gambar (yaitu warna, bentuk dan tekstur). Penggunaan centroid hasil pengelompokan dataset yang berasal dari RGB histogram dan matrik deteksi tepi dari beberapa gambar menggunakan FGKA, bisa digunakan sebagai acuan untuk melakukan pencarian. FGKA merupakan gabungan antara Algoritma Genetika dan Algoritma Kmeans. FGKA juga dikembangkan dari Genetic Kmeans Algorithm (GKA) yang selalu konvergen pada wilayah global. Pengelompokan dan pencarian gambar berdasarkan fitur warna-bentuk lebih baik dibandingkan berdasarkan fitur warna saja apabila menggunakan data-data yang dominan ke bentuk.

**Kata Kunci :** Algoritma Genetika, KMeans Clustering, CBIR, Pencarian Gambar

## I. Pendahuluan

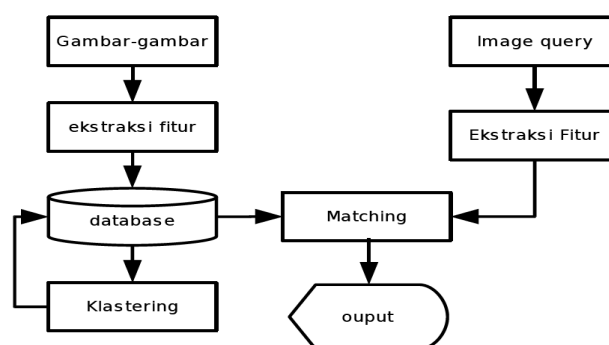
*Content-based image retrieval* (CBIR), yang juga dikenal dengan istilah *query by image content* (QBIC) dan *content-based visual information retrieval* (CBVIR) adalah suatu aplikasi computer vision yang digunakan untuk melakukan pencarian gambar-gambar digital pada suatu basis data.

Sekitar tahun 1970 penelitian awal *image retrieval* dilakukan dengan menggunakan pendekatan pengindeksan dan informasi citra berbasis teks. Teknik pencarian gambar dengan input teks dinilai tidak praktis karena dua alasan, yakni ukuran basis data yang besar dan subyektif dalam mengartikan gambar.

Untuk menghindari teknik tersebut, maka sekitar tahun 1990 *image retrieval* dikembangkan lagi menggunakan pendekatan alternatif yaitu teknik mencari gambar hanya berdasarkan informasi yang ada pada gambar tersebut. Teknik *image retrieval* yang dipilih disamping dapat mencapai rata-rata kemampuan retrieval yang tinggi, seringkali memberikan konsekuensi waktu komputasi yang tinggi dikarenakan harus memproses dimensi data gambar yang besar.

Proses secara umum dari CBIR adalah gambar yang menjadi *query* dilakukan proses ekstraksi fitur (*image contents*), begitu halnya dengan gambar yang ada pada sekumpulan gambar juga dilakukan proses seperti pada gambar *query*. Parameter fitur gambar yang dapat digunakan untuk retrieval pada sistem ini seperti histogram, susunan warna, tekstur, dan bentuk, tipe spesifik dari obyek, tipe event tertentu, nama individu, lokasi, emosi.

Beragam cara telah diajukan pada sistem CBIR ini, seperti pada penelitian sebelumnya yang mengelompokkan gambar berdasarkan fitur warna menggunakan FGKA (*Fast Genetic KMEANS Algorithm*) yang mana memiliki hasil yang cukup akurat. Sehingga, diharapkan dari penelitian selanjutnya akan lebih akurat apabila menggunakan beberapa fitur gambar lainnya.

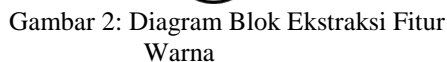


Gambar 1: Blok Diagram Sistem

## II. Metode Pencarian Gambar.

Ada tiga tahapan utama dalam pencarian gambar ini, yaitu Ekstraksi fitur, pengklasteran dan *matching* (pencocokan). Ekstraksi fitur adalah proses pengambilan histogram dan matriks deteksi tepi, baik dari gambar database maupun gambar query. Sedangkan *matching* (pencocokan), adalah proses perbandingan antara gambar query dengan gambar dalam database.

Seperti disebutkan diatas bahwa ekstraksi fitur warna adalah proses pengambilan histogram, baik dari gambar database maupun gambar query.



- Pengambilan nilai RGB tiap piksel.
- Kuantisasi warna dari yang semula berjumlah  $(360 \times 255 \times 255)$  atau 23409000 kemungkinan warna, diubah menjadi  $(4 \times 4 \times 4)$  atau 64 kemungkinan warna. Dengan cara ini, nilai R berkisar antara 0 sampai dengan 3, G berkisar antara 0 sampai dengan 3, dan B berkisar antara 0 sampai dengan 3.
- Normalisasi.
- Pembuatan histogram RGB, nilai-nilai ini nantinya dijadikan parameter ekstraksi fitur warna.

Gambar 3: Diagram Blok Ekstraksi Fitur Bentuk

- Citra asal dikonversi menjadi citra dengan derajat keabuan (gray-scale).  $S = (R + G + B) / 3$ .
- Sebelum melakukan proses deteksi tepi, konversi terlebih dahulu citra menjadi citra biner. Proses deteksi tepi (*Edge Detection*), menggunakan HPF (*High Pass Filter*), yakni proses filter yang mengambil citra dengan gradiasi intensitas yang tinggi dan perbedaan intensitas yang rendah akan dikurangi atau dibuang.

Sehingga hasilnya menjadi matriks berukuran 12 x 8 atau sebanyak 96 piksel yang berisi angka 0/1. Nilai-nilai inilah yang nantinya menjadi acuan dan merupakan hasil ekstraksi fitur bentuk.

Tahap ini merupakan implementasi dari algoritma FGKA untuk melakukan klasterisasi terhadap sejumlah hasil ekstraksi warna bentuk, sesuai dengan kedekatan jarak (kemiripan) antara gambar-gambar.

[illegible]

[illegible]

Contoh sederhana pada proses inisialisasi :

Sehingga FGKA akan mencari solusi terbaik hingga evolusi 10.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
9	5	7	7	0	3	2	4	1	6	6	8	7	0	9

Operator seleksi yang digunakan adalah seleksi proporsional. Hasil seleksi didapatkan dari populasi saat itu ( $S_1, S_2, \dots, S_Z$ ) yang mempunyai probabilitas ( $p_1, p_2, \dots, p_Z$ ) dengan definisi sebagai berikut:

Pada operator ini, tiap kromosom dikodekan dengan  $a_1 a_2 \dots a_N$  dan operator mutasi melakukan mutasi pada suatu gen  $a_n$  ( $n = 1 \dots N$ ) dengan nilai baru  $a_n'$  dengan sejumlah  $0 < MP < 1$  sebagai parameter yang dimasukkan oleh pengguna. Nilai tersebut dinamakan probabilitas mutasi. Mutasi dilakukan dengan  $a_n'$  yang dipilih secara random dari  $\{1, 2, \dots, K\}$  dengan distribusi  $(p_1, p_2, \dots, p_Z)$  yang didefinisikan dengan rumus :

Operator KMeans ini digunakan untuk mempercepat konvergensi. Solusi yang ada dikodekan dengan  $a_1 a_2 \dots a_N$ . Operator ini akan mengganti isi dari  $gen$   $a_n$  ( $n = 1 \dots N$ ) dengan nilai baru  $a_n'$ , dimana nilai yang baru merupakan klaster dengan jarak terdekat dari data  $a_n$  yang dihitung menggunakan rumus Euclidean. [10,14]

Berikut ini merupakan contoh hasil klastering dengan probabilitas mutasi (MP) = 0.1, nilai evolusi (NE) = 5, nilai populasi (NP) = 5.



Pada percobaan ini, gambar klaster 1 didominasi dengan gambar warna biru langit dan didominasi bentuk pesawat, dengan tingkat kemiripan 0.91 .

Dari segi fitur bentuknya, tidak terdapat kemiripan matriks deteksi tepi yang telah terkuantisasi antara gambar istana dan pesawat seperti terlihat pada gambar 9 dan 10.

[illegible]

1	1	1	1	1	1	1	1	1	1	1
1	0	0	0	1	0	0	1	1	0	0
1	0	1	1	0	0	0	1	0	1	0
1	0	1	1	0	0	0	0	0	1	0
1	1	0	0	1	1	1	1	1	0	0
1	0	0	0	0	0	0	0	0	1	0
0	0	0	0	1	0	1	1	1	0	0
1	1	1	1	0	1	1	1	1	1	1



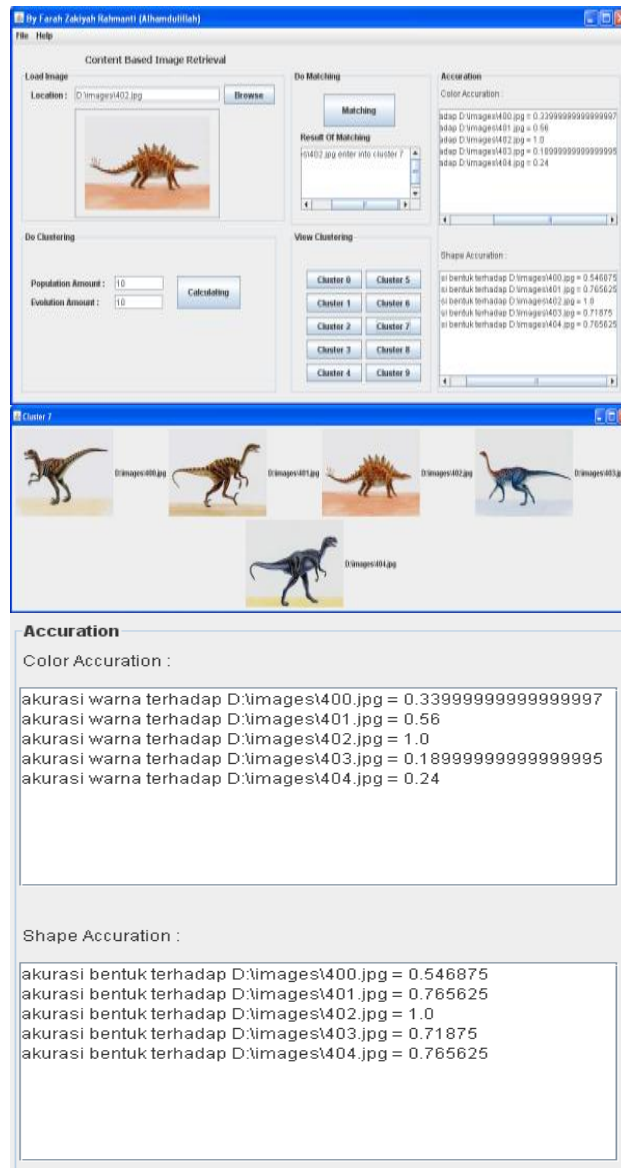
**Tabel 1: Hasil Pengklasteran (NE & NP 5 ; MP 0.1)**

Klaster	Tingkat Kemiripan
0	0.6875
1	0.91
2	0.8375
3	0.75
4	0.6
5	1
6	0.583
7	0.545
8	0.9
9	0.667
Rata-rata	0.748
TWCV	556480.35
Waktu Komputasi	3484

## Hasil Pencarian



Gambar 11: Contoh Gambar Query



Gambar 12: Contoh Hasil Pencarian

Gambar 12 menunjukkan hasil pencarian gambar menggunakan gambar query dinosaurus.

Tabel 2 : Hasil Pencarian (NE & NP 10 ; MP 0.1)

No	Item	Hasil
1	Jarak gambar query dan centroid terdekat	19.28730152198591
2	Waktu Komputasi	140
3	Rata-rata akurasi warna	0.46596
4	Rata-rata akurasi bentuk	0.759375

### Perbandingan Hasil Klustering

Tabel 3 : Perbandingan Hasil Klustering antara Fitur Warna-Bentuk, Warna, Bentuk

	Warna-Bentuk	Warna-Bentuk	Warna	Bentuk	Warna-Bentuk
NE & NP	5	10	10	10	10
Data training	1	1	1	1	2
Rata-rata kemiripan	0.76754	0.75805	0.80809	0.464054	0.842218

Berdasarkan tabel diatas, dapat disimpulkan :

- Klustering berdasarkan fitur bentuk memiliki tingkat kemiripan terendah, karena bentuk tidak dapat berdiri sendiri untuk gambar-gambar yang multiple fitur.
- Klustering berdasarkan fitur warna memiliki tingkat kemiripan yang lebih tinggi dibandingkan dengan fitur warna-bentuk bila menggunakan data training yang dominan ke warna..
- Klustering berdasarkan fitur warna memiliki tingkat kemiripan yang lebih rendah dibandingkan dengan fitur warna-bentuk bila menggunakan data training yang dominan ke bentuk.

### Perbandingan Hasil Pencarian

Tabel 4 : Perbandingan Hasil Pencarian antara Fitur Warna-Bentuk, Warna, Bentuk

	Warna-Bentuk	Warna-Bentuk	Bentuk	Warna-Bentuk
NE & NP	5	10	10	10
Data training	1	1	1	2
Rata-rata akurasi	0.45878	0.537	0.579142	0.43782

Berdasarkan tabel di atas, dapat disimpulkan bahwa :

Ketepatan pencarian antara gambar query dengan gambar yang ada dalam database memiliki rata-rata akurasi yang tidak lebih dari 0.6 . Karena hanya terdapat satu gambar di dalam database yang tepat dimana benar-benar sama dengan gambar query-nya.

Selain itu, gambar-gambar lainnya yang masuk dalam klaster yang sama, diasumsikan mirip berdasarkan content image-nya.

### Perbandingan Waktu Komputasi Hasil Pencarian

Tabel 5 : Perbandingan Waktu Komputasi Hasil Pencarian antara Fitur Warna-Bentuk, Warna, Bentuk

	Warna-Bentuk	Warna-Bentuk	Warna	Bentuk	Warna-Bentuk
NE & NP	5	10	10	10	10
Data training	1	1	1	1	2
Rata-rata waktu komputasi (ms)	196.8	206	175.2	580.8	240.6

Berdasarkan tabel di atas, dapat disimpulkan bahwa, proses pencarian berdasarkan fitur warna memiliki waktu komputasi yang lebih cepat dibandingkan pencarian berdasarkan warna-bentuk maupun bentuk, karena memiliki dimensi lebih sedikit (hanya 64) dari pada yang warna-bentuk & bentuk (160 atau 96).

### IV. Kesimpulan

Dari hasil pengujian dan analisa data yang telah dipaparkan tadi, dapat disimpulkan bahwa:

1. FGKA dapat digunakan pada proses pencarian gambar dengan terlebih dahulu mengelompokkan gambar-gambar yang memiliki nilai Histogram RGB dan matriks deteksi tepi yang berdekatan. Dengan cara ini, pada beberapa pengujian, hasil yang dikembalikan ternyata tidak tercampur.
2. Fitur warna-bentuk lebih sensitif dibandingkan dengan fitur warna, sehingga menyebabkan tingkat error yang lebih besar pada penelitian kali ini dibandingkan penelitian yang dulu pernah dilakukan apabila menggunakan data training yang dominan ke warna.
3. Fitur warna-bentuk lebih baik dibandingkan dengan fitur warna dalam proses pengklasteran apabila menggunakan data training yang dominan ke bentuk.

### V. Daftar Pustaka

- [1] Achmad Basuki, 2005, *Image Filtering*, Surabaya : Politeknik Elektronika Negeri Surabaya Institut Teknologi Sepuluh Nopember.
- [2] Achmad Basuki, Nana R, 2008, *Fitur Bentuk pada Citra*, Surabaya : PENS-ITS.
- [3] Acmad Basuki, 2003, *Suatu Alternatif Penyelesaian Permasalahan Optimasi dan Machine Learning Algoritma Genetika*, Surabaya : Politeknik Elektronika Negeri Surabaya Institut Teknologi Sepuluh Nopember.
- [4] Ali Ridho Barakbah, 2006, *Clustering*, Surabaya : Workshop Data Mining 18-20 Juli 2006 , Jurusan Teknologi Informasi, Politeknik Elektronika Negeri Surabaya.

- [5] Ali Ridho Barakbah, *Introduction to Machine Learning*, Surabaya : Politeknik Elektronika Negeri Surabaya Institut Teknologi Sepuluh Nopember.
- [6] Anonym, *Histogram Warna pada Image*, Surabaya : Politeknik Elektronika Negeri Surabaya Institut Teknologi Sepuluh Nopember.
- [7] Anonym, *Color Models and Color Applications*, Chapter 12.
- [8] Anonym, *Color histogram*,  
From [http://en.wikipedia.org/wiki/Color\\_histogram](http://en.wikipedia.org/wiki/Color_histogram) .  
5 Desember 2009.
- [9] Bayu Bagus, 2007, *Image Database Menggunakan Sistem Content Based Image Retrieval dengan Ekstraksi Fitur Terstruktur*, Surabaya : Jurusan Teknologi Informasi, Politeknik Elektronika Negeri Surabaya Institut Teknologi Sepuluh Nopember.
- [10] Entin Martiana, *Perbaikan Kinerja Algoritma Klusterisasi KMeans Genetika*, Surabaya : FTIF-ITS.
- [11] Fadlisyah, 2007, *Computer Vision dan Pengolahan Citra*, Jogjakarta : Penerbit Andi.
- [12] Nana Ramadijanti, 2008, *Image Processing*, Surabaya : Laboratorium Computer Vision, Politeknik Elektronika Negeri Surabaya Institut Teknologi Sepuluh Nopember.
- [13] Yanu Widodo, 2008, *Pencarian Gambar Berdasarkan Fitur Warna dengan GA-KMEANS Clustering*, Surabaya : Politeknik Elektronika Negeri Surabaya Institut Teknologi Sepuluh Nopember.
- [14] Yi Lu, Shiyong Lu, Farshad Fotouhi, 2004, *Poster Abstract FGKA (Fast Genetic KMEANS Algorithm)*, USA : Department of Computer Science Wayne State University Detroit, MI 48202.
- [15] Yue Zhang, 2002, *On the use of CBIR in Image Mosaic Generation*, Canada : Department of Computing Science, University of Alberta, Edmonton, Alberta.